



ENHANCING PHISHING DETECTION: A HYBRID DEEP LEARNING MODEL INTEGRATING BI-LSTM AND BI-GRU ALGORITHMS FOR URL AND CONTENT ANALYSIS

Sakthipriya N.¹, Dr. V. Govindasamy², Abhinesh C.³, Krishna Govarthini T⁴, Rajesh R.⁵, Swedha R.⁶

^{1, 2, 3, 4, 5, 6} Puducherry Technological University, Puducherry

ABSTRACT

In response to the pressing cybersecurity challenges posed by the proliferation of phishing URLs and malicious content, this research introduces a groundbreaking approach centered on transfer learning within deep neural networks. By leveraging transfer learning, intricate patterns within URLs and their content are unveiled, culminating in the development of a model seamlessly integrating Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) networks.

These architectures effectively capture sequential dependencies, enhanced by their bidirectional variants accessing both past and future states to comprehend temporal dynamics and improve performance. Through meticulous evaluation and fine-tuning processes, the proposed cybersecurity solution demonstrates robustness and efficacy in defending against evolving threats.

This research contributes significantly to advancing the cybersecurity domain, introducing an adaptive strategy that harnesses the strengths of Bi-LSTM and Bi-GRU networks within the framework of transfer learning, paving the way for more resilient and effective cybersecurity solutions.

KEYWORDS: Deep Learning Methods, Malware, Phishing URLs, Cybersecurity

INTRODUCTION

Malicious URLs and content present serious risks in the world of digital networks since they act as trickery access points for fraud, cyberattacks, and scams. These carefully crafted URLs have the potential to spread malware, start spear-phishing or phishing campaigns, and aid in other types of online fraud. Their threat stems from their propensity to blend in, which makes them difficult to spot and more likely to be ignored.

As the human factor in cybersecurity is acknowledged, education becomes essential. Users that receive security awareness training are better equipped to recognize and handle the complex web of harmful links. Organizations may improve their overall resistance against the ubiquitous threat of harmful URLs by cultivating a culture of cyber literacy and caution. This will make the digital world more secure for both individuals and enterprises.

Phishing connections represent yet another dishonest technique employed by cybercriminals to take advantage of people and institutions. Usually, the purpose of these links is to deceive users into giving sensitive information such as bank account data, login passwords, or personal information. They are placed inside emails, texts, or webpages that appear legitimate. Phishing connections frequently use social engineering techniques, in which hackers create websites or communications that look like trustworthy organizations in order to instill a false sense of urgency and trust.

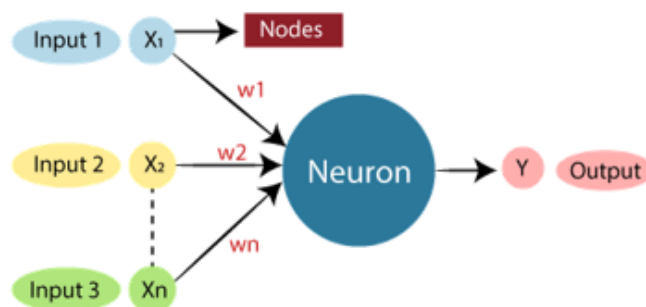


Fig 1 Deep learning architecture

Users should use caution and confirm the legitimacy of unexpected messages or emails before clicking on embedded links in order to combat phishing risks. To teach users how to spot and steer clear of phishing efforts, firms must implement email filtering systems and security awareness training. An additional line of protection against these misleading links comes from online browsers and cybersecurity software, which frequently include anti-phishing tools to identify and prevent access to known dangerous websites.

When it comes to cybersecurity, awareness, education, and cutting-edge technologies continue to be essential components of protecting against ever-changing dangers such as malicious contents and phishing attempts.

MATERIALS AND METHODS

This strategy departs from previous approaches in the realm

of cybersecurity by introducing a novel approach based on Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) networks.. These architectures are known for their proficiency in capturing sequential dependencies and understanding temporal dynamics within data, making them particularly suitable for the task of detecting phishing URLs and malicious contents. To optimize the accuracy and efficiency of phishing threat detection, the hybrid model incorporating Bi-LSTM and Bi-GRU employs both soft and hard voting mechanisms. Soft voting involves combining predicted probabilities from multiple models, while hard voting makes decisions based on the majority vote of the models. This ensemble technique ensures a more robust and well-rounded decision-making process, contributing to the overall performance enhancement of the system.

Crucially, the system is designed to simultaneously detect both phishing links and malicious contents, showcasing its capability to identify URLs exhibiting characteristics associated with phishing attacks and those linked to malicious activities. By addressing both aspects concurrently, the system provides a comprehensive defense mechanism, offering a more thorough and integrated approach to cybersecurity, particularly in combating evolving threats related to phishing and malicious links. This holistic strategy ensures that the system is adept at adapting to the dynamic nature of cyber threats in the digital landscape.

Bilstm (Bidirectional Long Short-Term Memory)

Recurrent neural network (RNN) architectures such as Bidirectional Long Short-Term Memory (BiLSTM) are frequently employed for sequence processing applications, including time series analysis and natural language processing. The main characteristic of a BiLSTM is that it is made up of two LSTM layers, one of which processes the input sequence forward and the other of which processes it backward. The input sequence is processed by the forward long short-term memory (LSTM) from beginning to finish, and by the backward LSTM from end to beginning in the opposite direction. The BiLSTM may gather data from both the input sequence's past and future states because to this bidirectional processing.

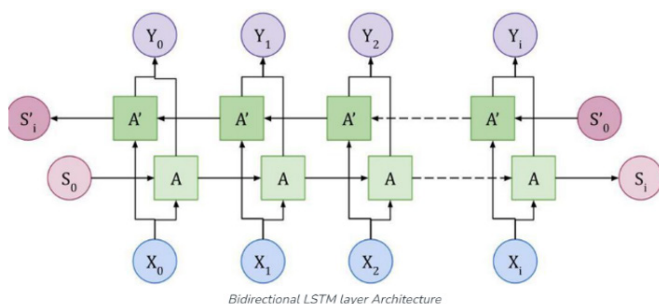


Fig 2 Bi-LSTM layer

By having two LSTM layers operating in opposite directions, a BiLSTM effectively increases the amount of context available to the network. For example, when processing a sentence, the forward LSTM can recognize each word's context based on the terms that come before it, while the backward LSTM

can understand the context based on the words that come after it. Combining information from both directions enables the BiLSTM to have a more comprehensive understanding of the input sequence.

Bigru (Bidirectional Gated Recurrent Unit)

Bi-GRU, short for bidirectional gated recurrent unit, represents a recurrent neural network architecture designed to effectively capture contextual information from input sequences. The Bi-GRU model, which consists of two independent GRU (Gated Recurrent Unit) layers, processes input data both forward and backward. Independently analyzing the sequence, each GRU layer makes use of its gating mechanisms to manage information flow and identify long-range dependencies in the data.

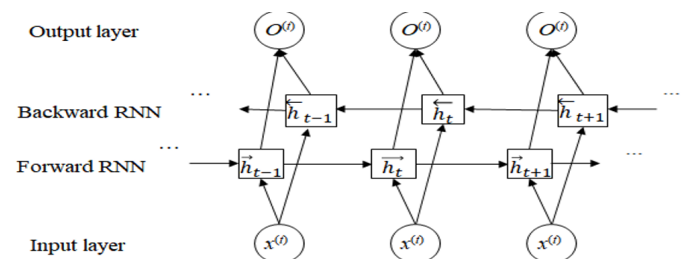


Fig 3 Bi-GRU layer

While the second GRU layer works in reverse, examining the input sequence from the beginning to the end, the first GRU layer examines the input sequence sequentially in the forward direction. Through bidirectional processing, the Bi-GRU model is better able to comprehend the temporal dynamics and linkages within the sequence by extracting contextual information from both past and future states of the input data.

Data Collection:

Data collected from Kaggle's open-source platform provides valuable datasets sourced globally, fostering innovation and collaboration. Researchers leverage Kaggle's diverse datasets for various projects including machine learning and data science.

Pre-processing:

Data undergoes cleaning and organization, handling missing values, duplicates, and outliers. Normalization and encoding standardize features and convert categorical variables for modelling.

Feature Extraction:

Extracted and encoded features from URLs and textual content. For URLs we parsed essential components such as scheme, domain, and path while computing features such as length and special character presence. Categorical features were numerically encoded using label encoding. Textual content underwent preprocessing steps including lowercase transformation, Regular Expression-based cleaning and application of NLTK tools for refinement such as removing stopwords and lemmatization. Finally, word embedding was employed to convert the cleaned text into numerical representations.

Model Creation using BI-LSTM and BI-GRU:

Bidirectional LSTM and Bidirectional GRU types of recurrent neural networks capture sequential patterns in data. Integrating these architectures allows the model to learn both forward and backward sequences ideal for analysing URLs and content for phishing detection.

Testing:

We tested our model's performance using a new dataset that it hadn't seen during training. This helped us ensure that our model could handle new situations effectively, just like it did during training. By evaluating its performance on unseen data, we could confirm that our model was reliable and ready for real-world use.

Prediction:

The trained model determines if a given instance signifies phishing by analyzing both URL and content features, and then compares its predictions to the ground truth labels to assess accuracy, precision, and recall. This comprehensive evaluation ensures the model's effectiveness in identifying potential phishing instances, providing valuable insights into its performance and reliability.

RESULTS AND DISCUSSION

Our project embarked on the development and implementation of a hybrid deep learning approach for phishing detection integrating both URL and content analysis. Initially, we trained separate models using Bi-LSTM, Bi-GRU and a hybrid Bi-LSTM and Bi-GRU architecture. Through rigorous experimentation our hybrid model consistently outperformed the individual models, demonstrating its efficacy in capturing nuanced patterns indicative of phishing attempts.

Upon successful validation of our hybrid model's superior performance, we proceeded to deploy it on a live website dedicated to detecting phishing attempts. Users were provided with a straightforward interface to input URLs, enabling real-time evaluation of their safety.

The implementation of our hybrid approach showcased remarkable accuracy in identifying potential phishing URLs, effectively empowering users to make informed decisions about their online safety. By leveraging the strengths of both Bi-LSTM and Bi-GRU algorithms, our model demonstrated robustness in scrutinizing both URL structures and content, thereby enhancing overall phishing detection capabilities.

This deployment underscores the practical relevance and effectiveness of our hybrid deep learning approach in real-world scenarios. However, while our results are promising, continuous refinement and augmentation of our model could further bolster its performance and contribute to the ongoing efforts in combating phishing attacks effectively.

Future iterations may involve exploring additional datasets, fine-tuning parameters and integrating advanced features to ensure continued efficacy and adaptability in the ever-evolving landscape of cybersecurity threats.

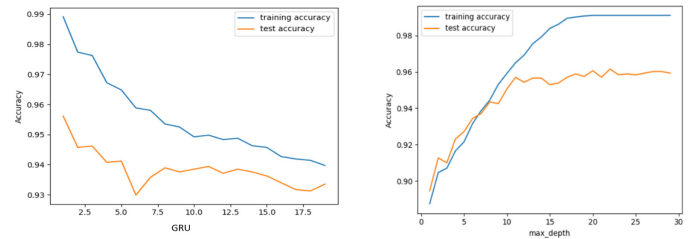


Fig 4 Accuracy graph

CONFUSION MATRIX

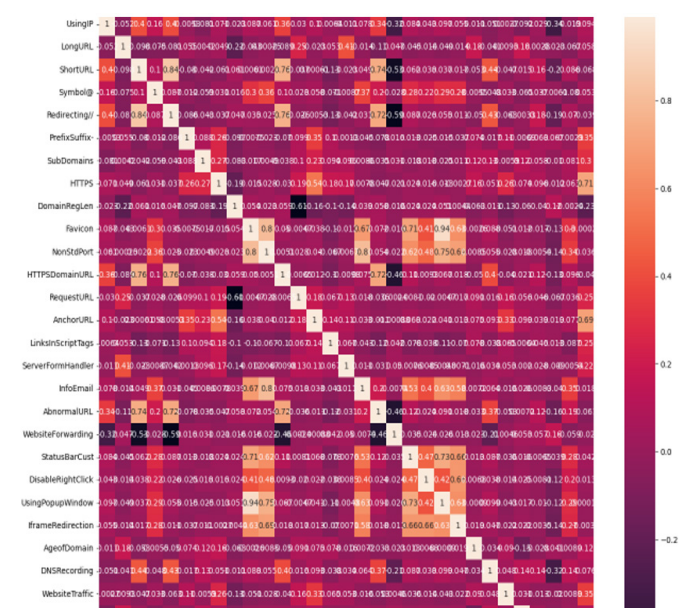


Fig 5 Confusion matrix

Representation: Presented as a confusion matrix.

Values: Each cell in the matrix contains the correlation coefficient between two features.

Interpretation: Positive correlation occurs when two features rise or fall together; negative correlation occurs when one feature rises while the other falls; and little to no link occurs when the values are close to zero.

Use: Helpful for identifying relationships and dependencies between features, assisting in feature selection, and understanding the multicollinearity (interdependence) within the dataset. It can generate a correlation matrix using libraries such as NumPy and pandas, and visualize it using tools like Seaborn or Matplotlib.

CONCLUSION

In conclusion, the proposed cybersecurity system represents a significant advancement in the field, leveraging cutting-edge techniques such as Bidirectional Gated Recurrent Unit (Bi-GRU) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks integrated through transfer learning. Through meticulous research and development, this system addresses the pervasive challenges associated with detecting and mitigating phishing URLs and malicious contents, which are critical vectors for cyber-attacks in today's digital landscape. By harnessing the power of deep learning and transfer learning, the system demonstrates robust performance in accurately identifying and classifying threats while minimizing false positives and

negatives. The comprehensive performance study, which employs stringent metrics including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC) curve, validates the efficacy and viability of the built models. Furthermore, the adaptability of the system to the dynamic nature of cyber threats is underscored by its ability to continuously learn and evolve, guided by insights gleaned from the performance analysis. This adaptability is crucial for staying ahead of emerging threats and ensuring the resilience of cybersecurity defenses in an ever-changing landscape. Overall, the proposed cybersecurity system represents a sophisticated and adaptive approach to combating cyber threats, offering a comprehensive defense mechanism against phishing URLs and malicious contents. As cyber threats continue to evolve, the system stands poised to provide robust protection, contributing to the ongoing efforts to safeguard digital assets and privacy in an increasingly interconnected world.

REFERENCES

1. Sophiya Shikalgar, S.D. Sawarkar, Swati Narwane; "Detection of URL based Phishing attacks using Machine Learning"; IJERT-International Journal of Engineering Research and Technology; 2019; Vol. 8; pp 537-544
2. Cho Do Xuan, Hoa Dinh Nguyen, Tisenko Victor Nikolaevich; "Malicious URL Detection Based on Machine Learning"; International Journal of Advanced Computer Science and Applications; 2020; Vol. 11; pp 148-153
3. Sanjiban Sekhar Roy, Ali Ismail Awad, Lamesgen Adugnaw Amare, Mabrie Tesfaye Erkihun, Mohd Anas; "Multimodal Phishing URL Detection using LSTM, Bidirectional LSTM and GRU Models"; 2022, Vol. 14; <https://doi.org/10.3390/fi14110340>
4. Subash Ariyadasa, Shantha Fernando, Subha Fernando; "Combining Long-Term Recurrent Convolutional and Graph Convolutional Networks to Detect Phishing sites using URL and HTML"; IEEE Access on Computer Science and Engineering; 2022; Vol. 10; DOI 10.1109/ACCESS.2022.3196018
5. Abdul Karim, Mobeen Shahroz, Khabib Mustofa, Samir Brahim, Belhaouari, S Ramana Kumar Joga; "Phishing Detection System Through Hybrid Machine Learning based on URL"; IEEE Access on Information and Computer Technology; 2023; Vol. 4; DOI 10.1109/ACCESS.2023.3252366.
6. Swatej Patil, Mayur Patil, Kotadi Chinnaiah; "Machine Learning and Deep Learning for Phishing Page Detection"; Universal Wiser International Conference on Machine Learning; 2023; DOI: <https://doi.org/10.37256/2320232629>; pp 45-54
7. Anjaneya Awasthi, Noopur Goel; "Phishing website prediction using base and ensemble classifier techniques with cross validation"; SpringerOpen – Cybersecurity; 2022; <https://doi.org/10.1186/s42400-022-00126-9>.
8. Doyen Sahoo, Chenghao lua, Steven C. H. Hoi, Malicious URL Detection using Machine Learning: A Survey, arXiv:1701.07179v3 [cs.LG], 21 Aug 2019
9. F. Vanhoenshoven, G. Napoles, R. Falcon, K. Vanhoof and M. K ' oppen, " Detecting malicious URLs using machine learning techniques, 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850079.
10. <https://www.kaggle.com/xwolf12/malicious-and-benign-websites> accessed on 27.01.2021
11. <https://openphish.com/> accessed on 27.01.2021